HPC ADVISORY COUNCIL

FLOW-3D Performance Benchmark and Profiling

September 2012

FLOW Science
Improving the world through accurate flow modeling

hp

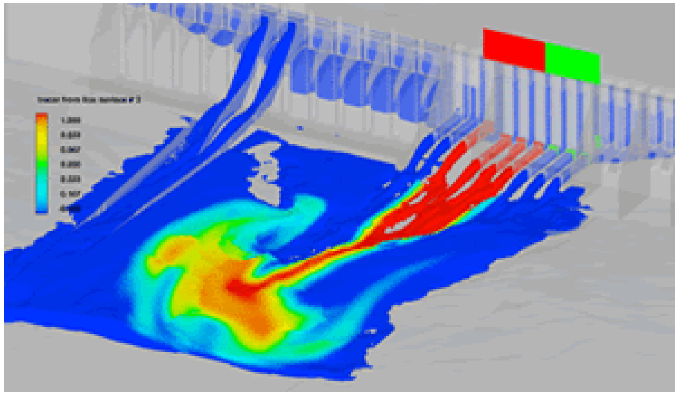
Mellanox TECHNOLOGIES

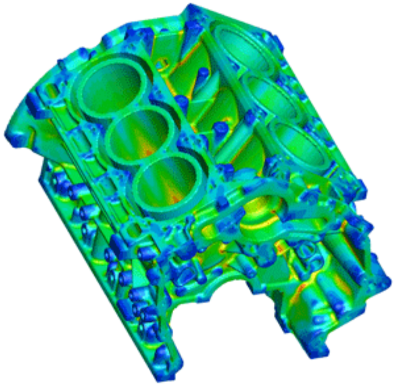
Note

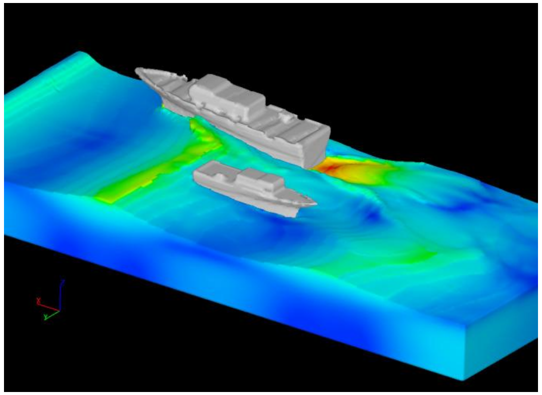
- **The following research was performed under the HPC Advisory Council activities**
  - Special thanks for: HP, Mellanox
- **For more information on the supporting vendors solutions please refer to:**
  - www.mellanox.com, http://www.hp.com/go/hpc
- **For more information on the application:**
  - http://www.flow3d.com/

FLOW-3D

- **FLOW-3D is a powerful and highly-accurate CFD software**
  - Provides engineers valuable insight into many physical flow processes
- **FLOW-3D is the ideal computational fluid dynamics software**
  - To use in the design phase as well as in improving production processes
  - Provides special capabilities for accurately predicting free-surface flows
- **FLOW-3D is a standalone, all-inclusive CFD package**
  - Includes an integrated GUI that ties components from problem setup to post-processing







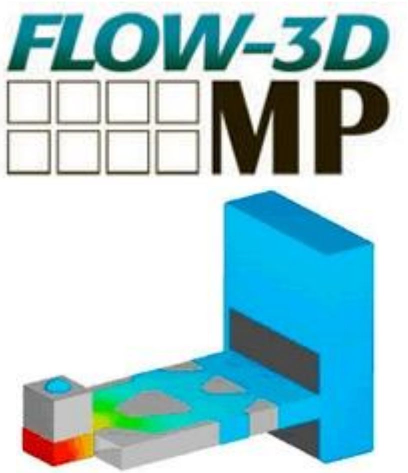
Objectives

- **The presented research was done to provide best practices**
  - FLOW-3D performance benchmarking
  - Interconnect performance comparisons
  - MPI performance comparison
  - Understanding FLOW-3D communication patterns

- **The presented results will demonstrate**
  - The scalability of the compute environment to provide nearly linear application scalability

Test Cluster Configuration

- **HP ProLiant SL230s Gen8 4-node “Athena” cluster**
  - Processors: Dual Eight-Core Intel Xeon E5-2680 @ 2.7 GHz
  - Memory: 32GB per node, 1600MHz DDR3 DIMMs
  - OS: RHEL 6 Update 2, OFED 1.5.3 InfiniBand SW stack
- **Mellanox ConnectX-3 VPI InfiniBand adapters**
- **Mellanox SwitchX SX6036 56Gb/s InfiniBand and 40G/s Ethernet Switch**
- **MPI: Intel MPI 4.0.3**
- **Application: FLOW-3D MP 4.2**
- **Benchmark Workload:**
- **Input dataset:**
  - P1_Inkjet

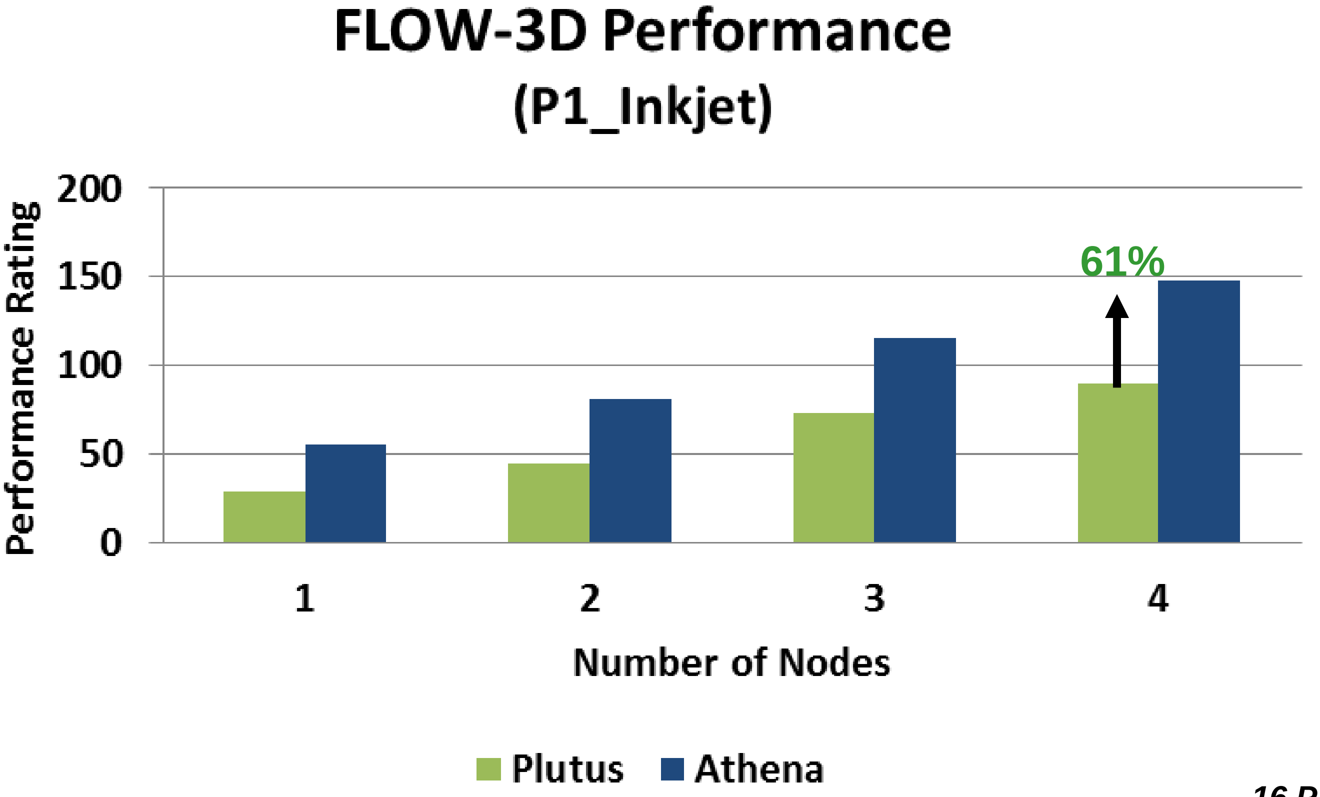


About HP ProLiant SL230s Gen8

Item	SL230 Gen8
Processor	Two Intel® Xeon® E5-2600 Series, 4/6/8 Cores,
Chipset	Intel® Sandy Bridge EP Socket-R
Memory	(512 GB), 16 sockets, DDR3 up to 1600MHz, ECC
Max Memory	512 GB
Internal Storage	Two LFF non-hot plug SAS, SATA bays or Four SFF non-hot plug SAS, SATA, SSD bays Two Hot Plug SFF Drives (Option)
Max Internal Storage	8TB
Networking	Dual port 1GbE NIC/ Single 10G Nic
I/O Slots	One PCIe Gen3 x16 LP slot 1Gb and 10Gb Ethernet, IB, and FlexF abric options
Ports	Front: (1) Management, (2) 1GbE, (1) Serial, (1) S.U.V port, (2) PCIe, and Internal Micro SD card & Active Health
Power Supplies	750, 1200W (92% or 94%), high power chassis
Integrated Management	iLO4 hardware-based power capping via SL Advanced Power Manager
Additional Features	Shared Power & Cooling and up to 8 nodes per 4U chassis, single GPU support, Fusion I/O support
Form Factor	16P/8GPUs/4U chassis

FLOW-3D Performance - Processors

- **Intel E5-2680 processors (Sandy Bridge) cluster outperforms prior CPU generation**
  - Performs 61% higher than X5670 cluster at 16 nodes
- **System components used:**
  - Athena: 2-socket Intel E5-2680 @ 2.7GHz, 1600MHz DIMMs, FDR InfiniBand, 1HDD
  - Plutus: 2-socket Intel X5670 @ 2.93GHz, 1333MHz DIMMs, QDR InfiniBand, 1HDD

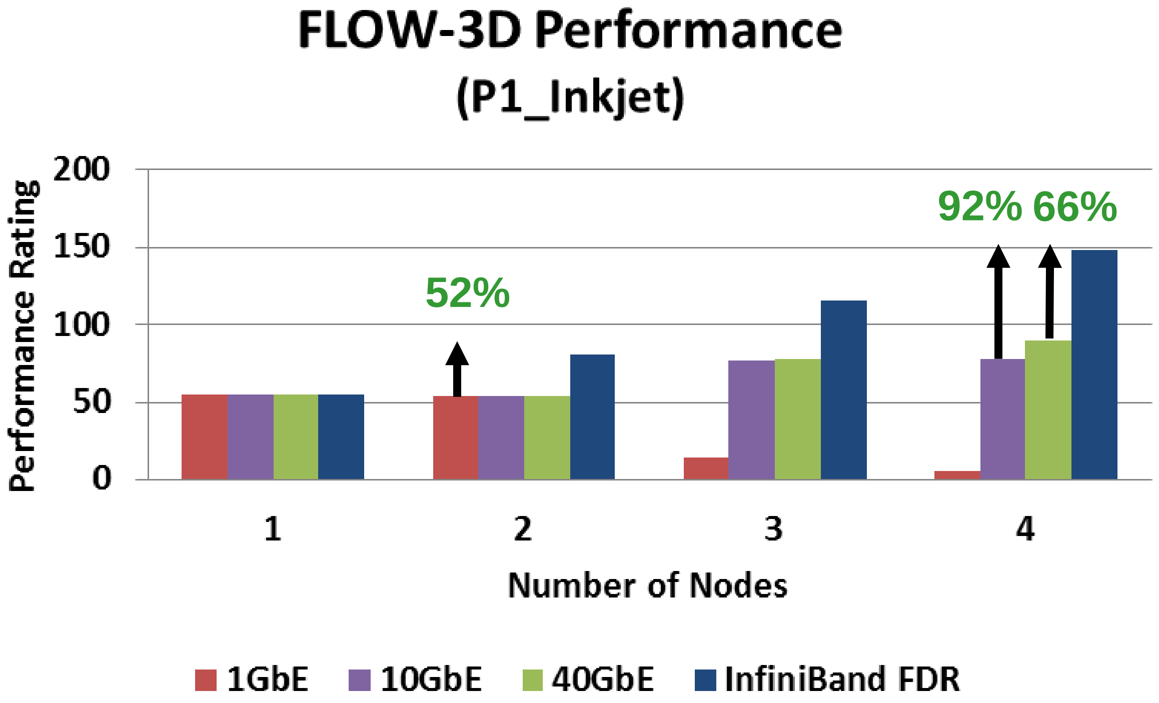


Higher is better

16 Processes/Node

FLOW-3D Performance - Interconnect

- **InfiniBand FDR is the most efficient inter-node communication for FLOW-3D**
  - Outperforms 10GbE by 92% at 4 nodes
  - Outperforms 40GbE by 66% at 4 nodes
  - Outperforms 1GbE by 52% at 2 nodes
- **1GbE do not show performance gain beyond 1 node**

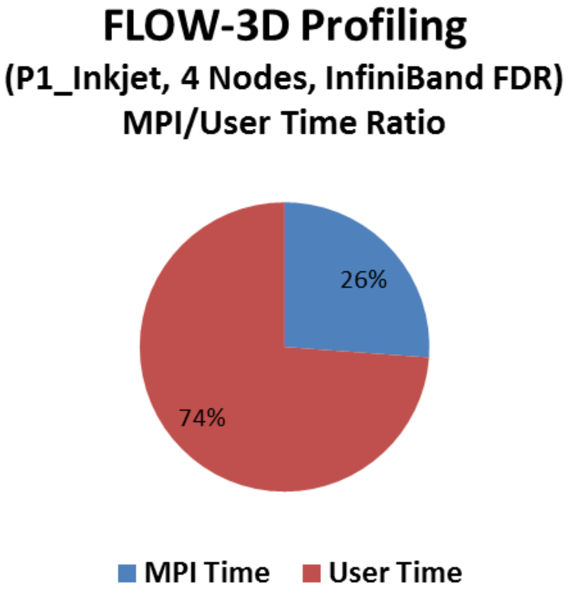


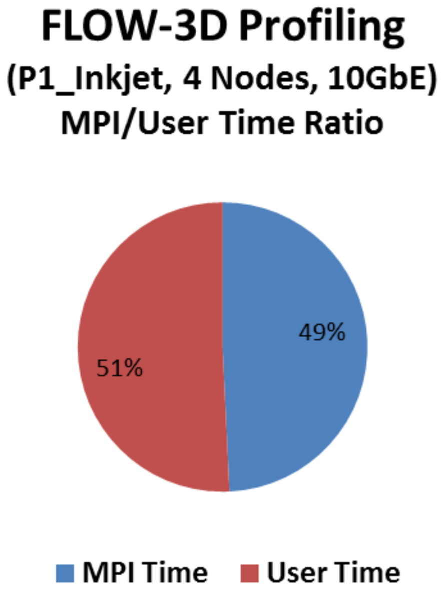
Higher is better

16 Processes/Node

FLOW-3D Profiling – MPI Time Ratio

- **InfiniBand FDR reduces the communication time at scale**
  - InfiniBand FDR consumes about 26% of total runtime
  - 10GbE consumes about half of total runtime

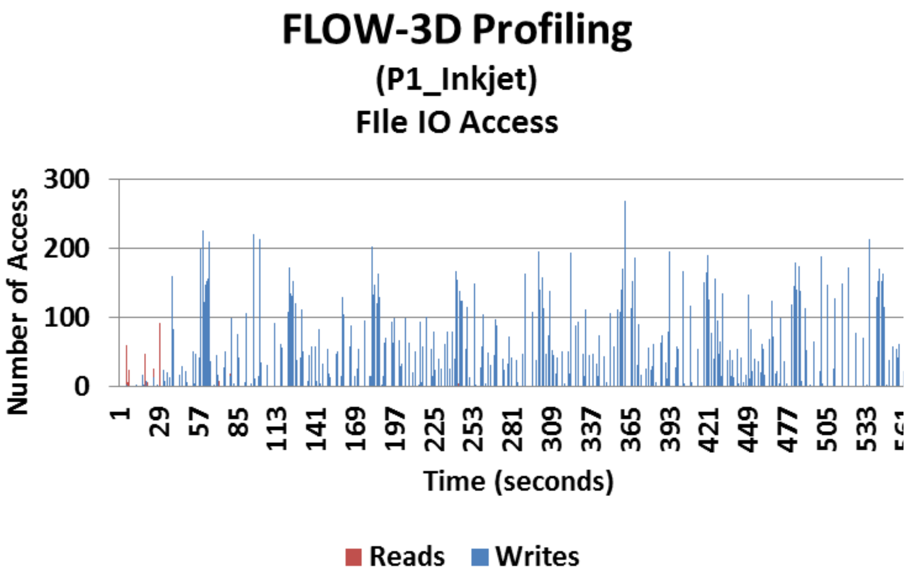


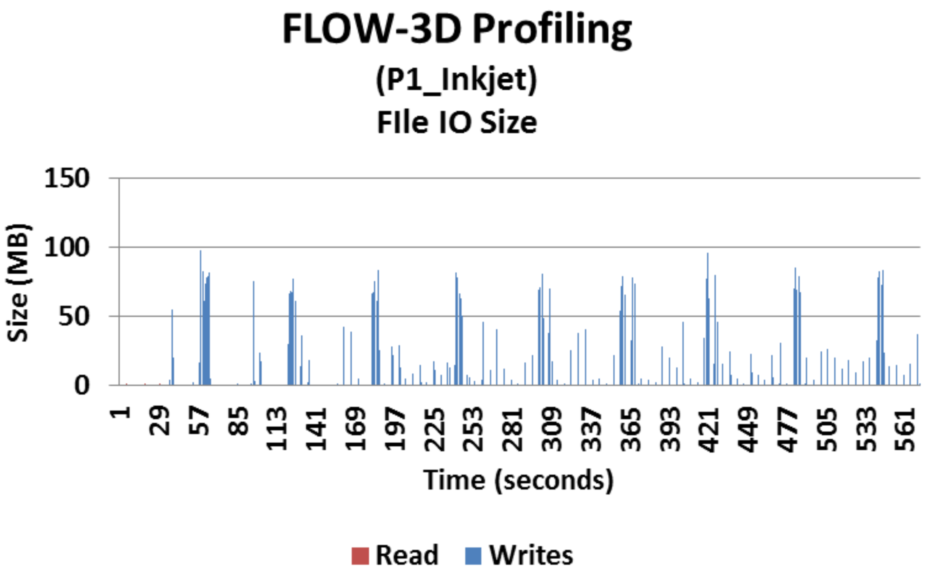


16 Processes/Node

FLOW-3D Profiling – Disk IO

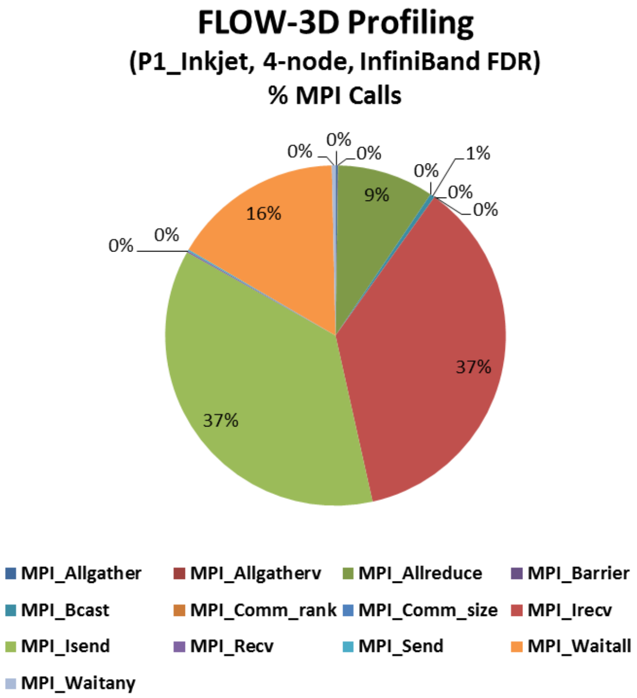
- **Heavy disk write access is seen throughout the test run**
  - Not much access for disk IO reads
  - Tests shows that FLOW-3D could benefit from better disk IO

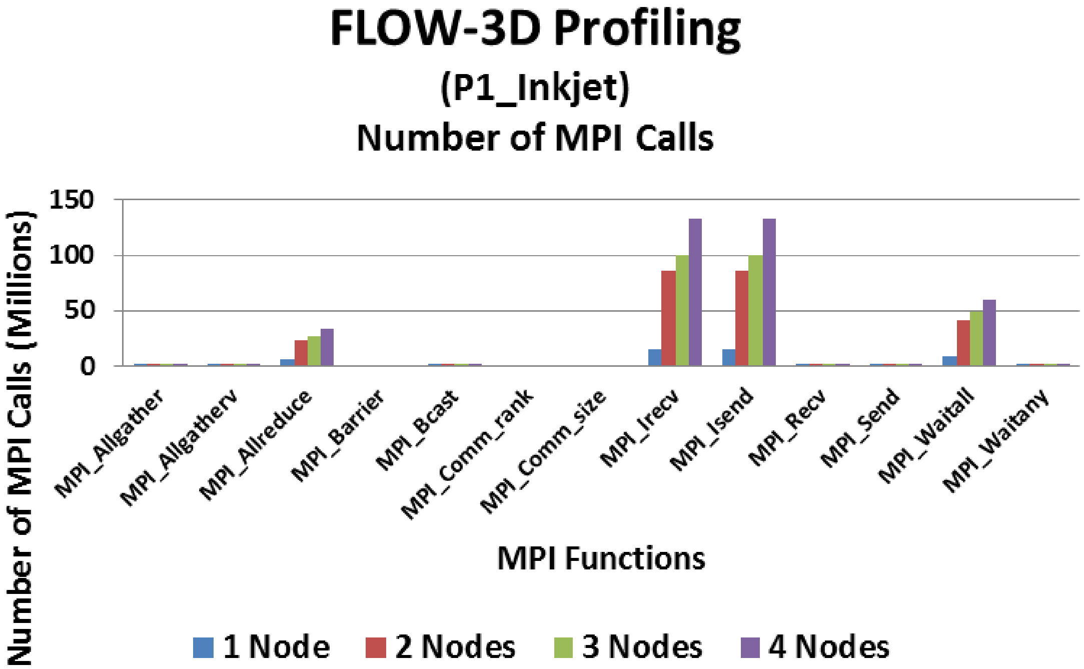




FLOW-3D Profiling – MPI Functions

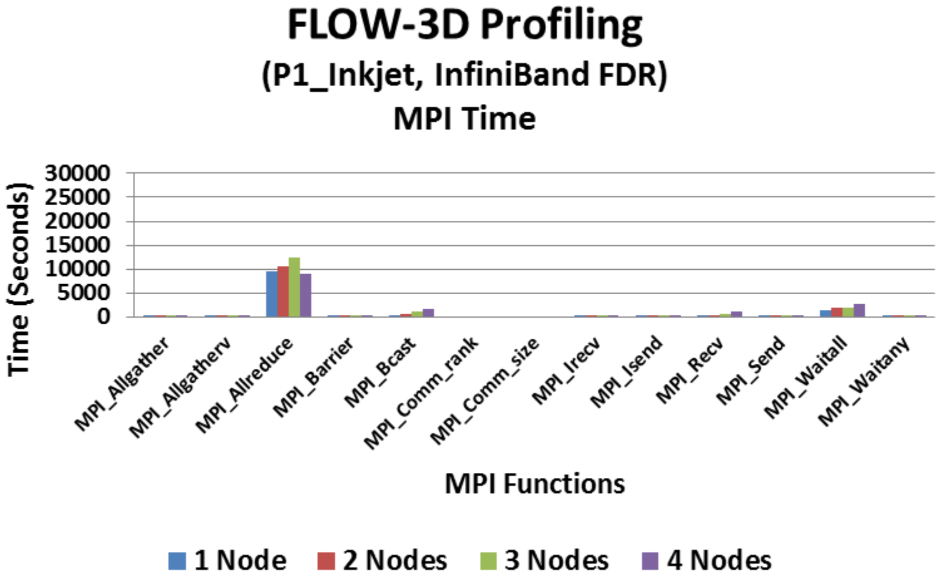
- **Mostly used MPI functions**
  - 4 nodes: MPI_Irecv (37%) and MPI_Isend (37%), MPI_Waitall (16%), MPI_Allreduce (9%)

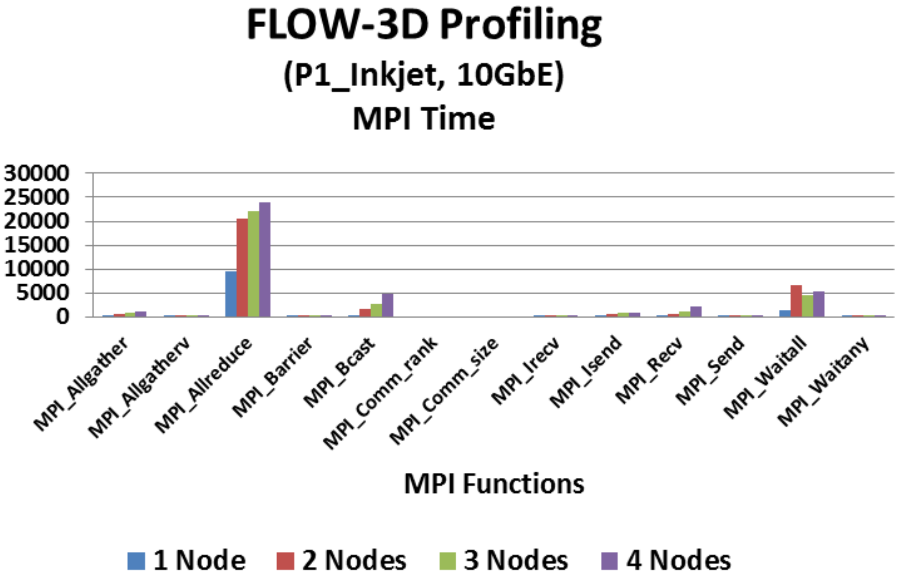




FLOW-3D Profiling – MPI Functions

- **The most time consuming MPI functions:**
  - InfiniBand FDR: MPI_Allreduce (60%), MPI_Waitall (18%), MPI_Bcast (11%)
  - 10GbE: MPI_Allreduce (62%), MPI_Waitall (14%), MPI_Bcast (13%)

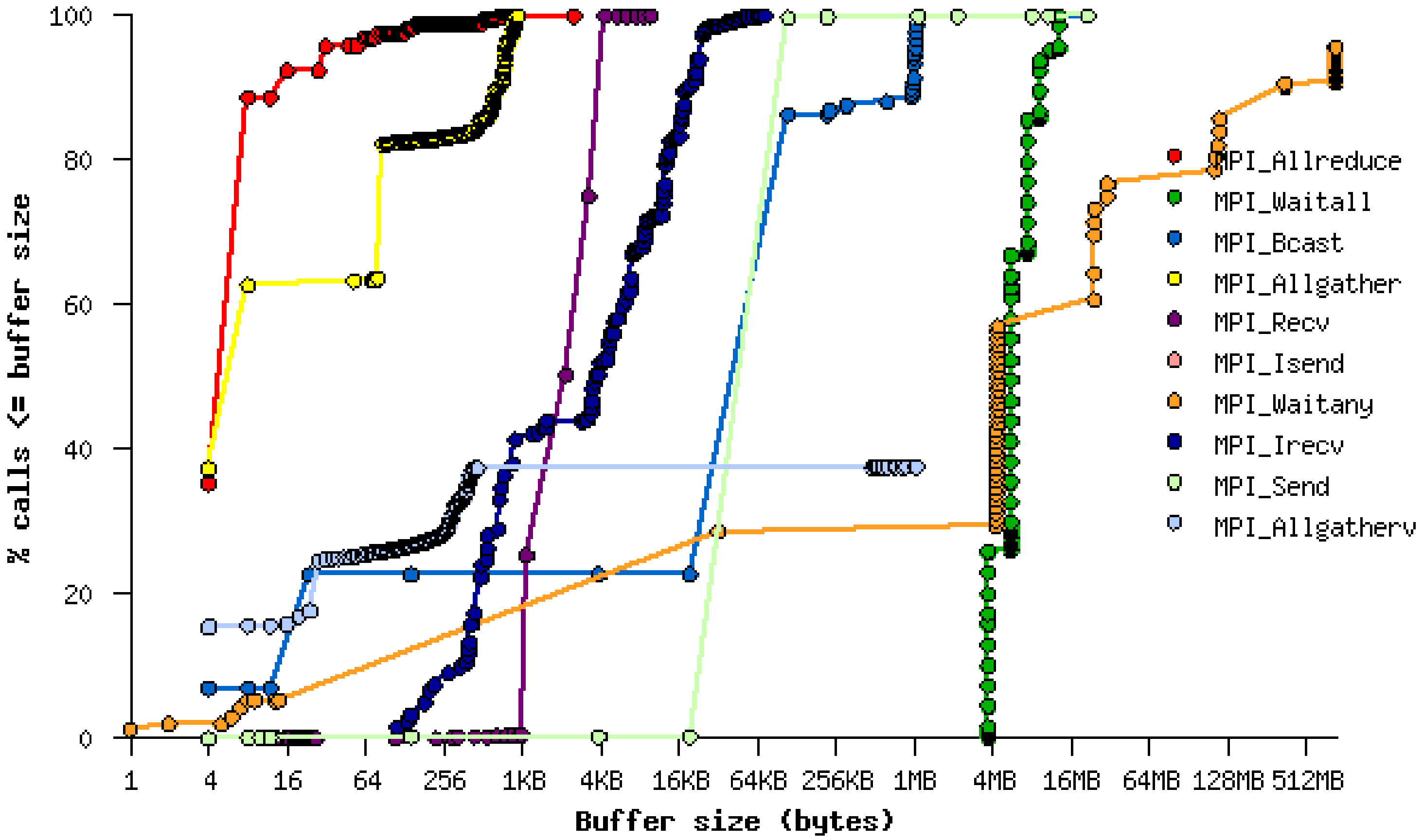




FLOW-3D Profiling – Message Size

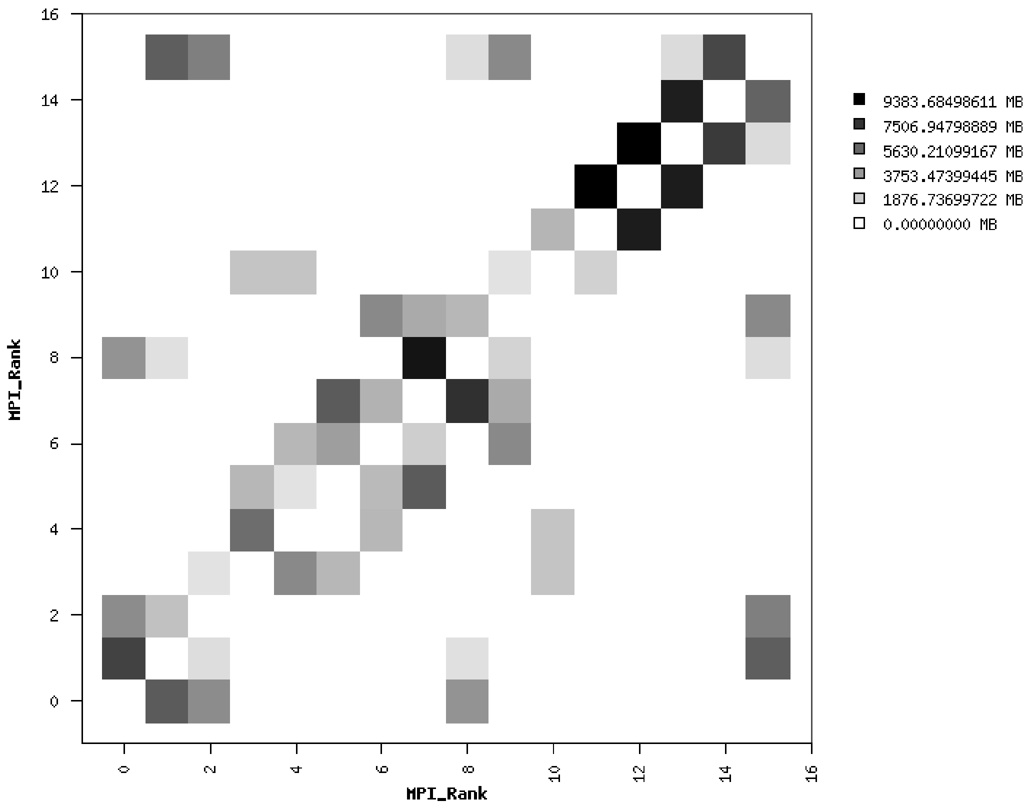
- **Distribution of message sizes for the MPI calls**
  - MPI_Irecv between 64B to 256KB
  - MPI_Allreduce: small messages from 4B to 1KB

P1_Inkjet



FLOW-3D Profiling – Point-to-point Flow

- **Heavy MPI communications seen between prorcesses**
  - Mainly concentrated between close neighboring ranks



P1_Inkjet

FLOW-3D Summary

- **HP ProLiant Gen8 servers delivers better FLOW-3D Performance than its predecessor**
  - ProLiant Gen8 equipped with Intel E5 series processes and InfiniBand FDR
  - Provides 61% higher performance than the ProLiant G7 servers when compare at 4 nodes
- **InfiniBand FDR is the most efficient inter-node communication for FLOW-3D**
  - Outperforms 10GbE by 92% at 4 nodes
  - Outperforms 40GbE by 66% at 4 nodes
  - Outperforms 1GbE by 52% at 2 nodes
- **FLOW-3D Profiling**
  - Heavy file IO writes are seen throughput the job run
  - Heavy MPI communications are seen between MPI processes
  - InfiniBand FDR reduces communication time; leave more time for computation
    - InfiniBand FDR consumes 26% of total time, versus 49% 10GbE
  - Non-blocking communications are seen:
    - MPI_Irecv (37%) and MPI_Isend (37%), MPI_Waitall (16%), MPI_Allreduce (9%)

Thank You

HPC Advisory Council

All trademarks are property of their respective owners. All information is provided "As-Is" without any kind of warranty. The HPC Advisory Council makes no representation to the accuracy and completeness of the information contained herein. HPC Advisory Council Mellanox undertakes no duty and assumes no obligation to update or correct any information presented herein